Training: Google Cloud
Generative AI in Production

Google Cloud

TRAINING GOALS:

Traditional MLOps is a set of practices to productionize traditional ML systems for enterprise applications. Generative AI raises new challenges in managing and productionizing applications at scale. The field of generative AI operations seeks to address these new challenges. In this course, you learn about the challenges that arise when deploying and productionizing generative AI-powered applications. You learn how to secure your generative AI-powered applications. Finally, you will discuss best practices for logging and monitoring your generative AI-powered applications in production.

What you'll learn:

- Understand the challenges in productionizing applications using generative AI
- Manage experimentation and evaluation for LLM-powered application
- Productionize LLM-powered applications
- Secure generative AI applications
- Implement logging and monitoring for LLM-powered applications

Audience

Developers, DevOps engineers and machine learning engineers who wish to operationalize GenAI-based applications.

Products

- Vertex AI

CONSPECT:

- Module 1 - Introduction to Generative AI in Production
  - Topics:
    - Generative AI operations
    - Traditional MLOps vs. GenAIOps

Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70
BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007

        - Components of an LLM system
        - RAG/ReAct architecture
    - Objectives:
        - Understand generative AI operations
        - Compare traditional MLOps and GenAIOps
        - Analyze the components of an LLM system
        - Define and compare RAG and ReAct
- Module 2 - Generative AI Application Deployment
    - Topics:
        - Application deployment options
        - Deployment, packaging, and versioning
    - Objectives:
        - Evaluate application deployment options
        - Deploy, package, and version apps
    - Activities:
        - Lab: Deploying an Agentic Application on Cloud Run
- Module 3 - Productionizing Generative AI
    - Topics:
        - Maintenance and updates
        - Testing and evaluation
        - CI/CD pipelines for gen AI-powered apps
    - Objectives:
        - Maintain and update LLM models
        - Test and evaluate gen AI-powered apps
        - Deploy CI/CD pipelines for gen AI-powered apps
    - Activities:
        - Lab: Tracking Versions of Generative AI Applications
- Module 4 - Securing Generative AI Applications
    - Topics:
        - Security challenges
        - Prompt security
        - Sensitive Data Protection and DLP API
        - Model Armor
    - Objectives:
        - Identify security challenges for gen AI applications
        - Understand prompt security issues

Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70
BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007

        - Apply sensitive data protection and DLP API
        - Implement Model Armor
    - Activities:
        - Lab: Securing Generative AI-Powered Applications
- Module 5 - Observability for Production LLM Systems
    - Topics:
        - Cloud Operations
        - Cloud Logging
        - Monitoring
        - Cloud Trace
        - Agent Analytics and AgentOps
        - Putting it all together
    - Objectives:
        - Describe the purpose and capabilities of Google Cloud Observability
        - Explain the purpose of Cloud Monitoring
        - Explain the purpose of Cloud Logging
        - Explain the purpose of Cloud Trace
    - Activities:
        - Lab: Logging, Monitoring, and Agent Analytics

REQUIREMENTS:

Completion of the "Application Development with LLMs on Google Cloud" or equivalent knowledge.

Difficulty level



CERTIFICATE:

The participants will obtain certificates signed by Google Cloud

TRAINER:

Authorized Google Cloud Trainer

Compendium Centrum Edukacyjne Spółka z o.o.
ul. Tatarska 5, 30-103 Kraków, tel.: (12) 298 47 77
ul. Postępu 18B, 02-676 Warszawa, tel.: (22) 417 41 70
BNP Paribas Bank Polska S.A.
EUR: PL 79 1600 1462 1853 6451 3000 0006
USD: PL 52 1600 1462 1853 6451 3000 0007